The roles of exposure and speed in road safety analysis

Xin Pei, S.C. Wong, N.N. Sze

Abstract

Speed is a determining factor in road safety analysis. It is generally believed that an increase in speed harms road safety. However, it can also be argued that driving at high speed reduces the length of time exposure and thus the likelihood of a crash. It is therefore critical to clarify the roles that exposure and speed play in road safety analysis. This study evaluates the relationship between speed and crash risk with respect to distance and time exposure, using disaggregated crash and speed data collected from 112 road segments in Hong Kong. A joint probability model based on a full Bayesian method is applied simultaneously to model crash occurrence and crash severity. In addition, we consider the explanatory variables, including road design, weather conditions, and temporal distribution, in the proposed crash prediction model. The results indicate that average speed plays a significant role in crash risk, despite opposing correlations with respect to distance and time exposure; the correlation between speed and crash risk is positive when distance exposure is considered, but negative when time exposure is used. However, in both cases, speed is positively associated with the injury severity.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In transportation engineering, speed is one of the basic variables that index the state and the level of service of roadway facilities. Speed plays an important role in traffic management and control. It is also generally believed to be a determining factor in the number and severity of crashes (OECD, 1996; National Research Council, 2010), because driving becomes a more complex and demanding task at high speed and thus the likelihood of driver error is greater. Furthermore, the energy released in a high-speed crash is greater, thereby increasing the likelihood of serious injury, death, and significant damage to property.

Average speed and speed dispersion both influence crash occurrence and crash severity (Hauer, 2009; National Research Council, 2010). Yet there is no consensus on the magnitude and direction of the influences of speed (Lave and Elias, 1994; Garber and Gadirau, 1988; Shinar, 1998; McCarthy, 1998; Taylor et al., 2000, 2002; Aarts and van Schagen, 2006) and speed dispersion (Lave, 1985; Zlatoper, 1991) on the risk of crash. Some studies have suggested that the direction of the influences of speed and speed dispersion on crash risk change when speed is increased to a certain level (Solomon, 1964). However, Nilsson (2004) used controlled before-and-after studies to establish the Power Model to describe the effect of changes in speed on changes in crash frequency and severity. The Power Model was validated by Elvik et al. (2004) using the meta-analysis approach. Empirical estimates of the exponents of the Power Model have been helpful for extended investigations into the association between speed and crash risk. A recent study suggested that the exponents of the Power Model might vary with road type. In particular, the exponent estimates for the risk of a slight injury crash were smaller than predicted by the Power Model on urban arterial roads (Cameron and Elvik, 2010). It seems that the magnitude of the effect of speed on road safety in urban road network needs to be further explored.

The availability of disaggregated speed data is essential for the analysis of the relationship between speed and road safety. With the support of advanced intelligent transport system (ITS) techniques, it is possible to detect traffic crashes based on comprehensive real-time speed and flow data obtained from a double-loop detector (Lee et al., 2002; Abdel-Aty et al., 2004, 2008). In this case, speed data are used as a proxy for roadway performance in acute traffic control and management, but not for inferences on the effect of speed on traffic safety. Disaggregated speed data are also essential for estimating length of time exposure in models of crash risk. Furthermore, when modeling crash risk, the explanatory factors such as road design, weather conditions, and temporal distribution for the association between speed, speed dispersion, and traffic safety must be considered.
Exposure measures the likelihood of being involved in a dangerous or hazardous situation. In road safety research, exposure is an important factor in the estimation of crash risk (Chapman, 1973; Wolfe, 1982; Rumar, 2002). The selection of an appropriate proxy for exposure is thus essential to evaluate the influences of possible risk factors on the crash risk. A wide range of data are used to proxy exposure, including traffic volume (Maou, 1994; Mountain et al., 1996; Qin et al., 2004, 2006; Wong et al., 2007; Van den Bossche et al., 2005), conflicts (Bie et al., 2005; Wong et al., 2006), travel distance (Li et al., 2003), travel time (Chipman et al., 1993), and population or fuel consumption (Fridstrøm et al., 1995).

In prior studies, distance exposure, measured in either vehicle kilometers (VKM) or vehicle miles traveled (VMT), was adopted to model the relationship between speed and crash risk (Abbas, 2004; Hakim et al., 1991). However, it could be argued that as length of time exposure on the road might be reduced when speed is increased, then the likelihood of crash involvement might also be reduced (Chipman et al., 1992). For this reason, time exposure measured in vehicle hours (VH) could be considered a better proxy for exposure. In this study, taking advantage of the availability of disaggregated speed data, it is possible to determine the length of time exposure given that the traffic volume and length of road segment are known. Hence, the influence of both exposure measures (i.e., distance and time exposure) on the relationship between speed and road safety can be evaluated.

In this study, comprehensive information on crashes, traffic flow, and speed were collected from 112 roadway segments in Hong Kong during the three-month period from July to September 2009. A joint probability model based on the MCMC approach full Bayesian method is established to model crash occurrence and crash severity simultaneously (Pei et al., 2011). The rest of this paper is organized as follows. In the following section, we first describe the study design and the method used for collecting data. We discuss the methods of analysis employed in this study in Section 3. The results are presented in Section 4 and their implications are discussed in Section 5. Section 6 presents our conclusions and recommendations for future research.

2. Data

In preparation for this study, we first established a comprehensive crash database containing information on traffic flow, speed, road design, weather conditions, and temporal distribution for roadway segments in Hong Kong, using geographical information system (GIS) techniques. Detailed traffic flow data were obtained from the annual traffic census (ATC) system that consists of over 1500 stations and covers 96.8% of all motorways in Hong Kong (Tong et al., 2003; Lam et al., 2003). The ATC system provides detailed information on average annual daily traffic (AADT) and its temporal (with respect to month, day of the week, and hour of the day) and directional multiplicative factors. In particular, continuous measurements are taken at 112 core stations. These core stations are evenly and widely distributed across the whole territory and cover a total length of 164.6 km (i.e., 8.0% of all motorways in Hong Kong). In the proposed model, we derived directional traffic volumes for all road segments adjacent to the 112 core stations for every 4-h period [07:00–11:00 (morning), 11:00–15:00 (noon), 15:00–19:00 (afternoon), 19:00–23:00 (evening), 23:00–03:00 (middle of the night) and 03:00–07:00 (dawn)] between July and September 2009. The sample thus comprises a total of 68,990 observations.

Speed data measured at the relevant road segments during the corresponding observation periods were obtained from measurements taken by 480 taxis equipped with global positioning system (GPS) devices operating on the roadway network in Hong Kong. In particular, instantaneous information on the location, speed, and travel direction of each GPS probe was transmitted to and stored in the control center at 30-s intervals. The average travel speed and the standard deviation of travel speed for every 4-h period can be calculated using these raw data. Here, average speed should account for traffic conditions, while the standard deviation of speed should account for variation in traffic conditions across the spatial–temporal domain. Hence, the speed data collected from the taxis can be used as a proxy for actual speed in particular traffic conditions and can facilitate analysis of the role that speed plays in road safety. When a crash occurs, the traffic at the road segment in question will be disrupted, thereby affecting traffic speed. In this case, the estimated speed is estimated on the basis of road design, weather conditions, and temporal distribution when there is no crash, rather than the speed directly observed during the crash period.

Crash data were obtained from the Traffic Information System (TIS) maintained by the Transport Department. This database contains precise information on crash, vehicle, and casualty attributes for all crashes involving personal injury that are reported to the police. We applied GIS techniques to map crashes that occurred on the studied road segments. A total of 347 crashes occurred at the corresponding road segments during the three-month period from July to September 2009. In the TIS, crashes are classified into three categories—fatal, serious, and slight—according to the severity of injuries among the casualties. In this study, we grouped the fatal and serious crashes together as killed and seriously injured (KSI) crashes. Of the 347 crashes captured, 59 (17.0%) were KSI.

This study considers both distance and time exposure. Distance exposure was derived by multiplying traffic volume by road segment length (in vehicle-kilometers, VKM). Given the average speed and road segment length, it was possible to estimate the travel time, and time exposure was then derived by multiplying traffic volume by average travel time (in vehicle-hours, VH). Information was also collected on the geometric design of the road segments and traffic control at each segment. The variables of interest are lane changing opportunity (LCO, which refers to the total number of possible lane-cuttings based on those set out by different lane markings, as shown in Fig. 1), road curvature (average changing angle), gradient, number of merging ramps, number of diverging ramps, number of intersections (including signal junctions, yield junctions and stop sign junctions), the presence of a central divider, the presence of a hard shoulder, the presence of a bus stop, and the presence of on-street parking. In addition, rainfall is also an important environmental factor likely to have a significant impact on road safety due to its effects on driver visibility and vehicle braking performance. In this study, detailed rainfall data for each geographical location and period were obtained from the Hong Kong Observatory. The influences of temporal distribution on crash risk, in terms of day of the week and time of day, are also controlled for. Table 1 summarizes the characteristics of the 68,990 observations analyzed in the proposed model.

A multicollinearity test for the independent variables was conducted prior to the determination of association measures for crash occurrence and crash severity. In the multicollinearity test, if the VIF value of any independent variable is greater than 10, that variable should be removed from the model to avoid biased parameter estimates. The results of the multicollinearity test revealed that the VIF values of the independent variables were all less than 10. Therefore, no evidence could be established for the existence of a multicollinearity problem.
Fig. 1. Lane changing opportunities for different road section configurations. Notes: A = double continuous lines. No lane changing is allowed between Lane 1 and Lane 2: \( LCO_A = 0 \); B = continuous line and broken line. Lane changing from Lane 3 to Lane 2 is allowed, while that from Lane 2 to Lane 3 is not allowed: \( LCO_B = 1 \); and C = single broken line. Lane changing between Lane 3 and Lane 4 is allowed: \( LCO_C = 2 \). Therefore, total \( LCO = LCO_A + LCO_B + LCO_C = 3 \).

### 3. Method

#### 3.1. Modeling framework

A joint probability model is established to model the crash occurrence and severity simultaneously using the MCMC approach full Bayesian method (Pei et al., 2011).

For crash occurrence, because the dependent variable—crash frequency—is discrete, non-negative, and random in nature, count data models are applied. In particular, to analyze the effects of distance and time exposure, logarithmically transformed VKM and VH are incorporated into the model to offset these respective effects.

As crash occurrence is subject to over-dispersion, a negative binomial regression model should be applied (Washington...
Consequently, the probability function is defined by
\[
P(y_i) = \frac{\Gamma[(1/\alpha) + y_i]}{\Gamma(1/\alpha) y_i!} \left( \frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left( \frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{y_i}
\]  
(1)
along with the link function, \( \lambda_i = \exp(\log(\text{exposure}) + \theta X_i) \), where \( y_i \) is the number of crashes for the \( i \)th observation and \( \lambda_i \) is the expected number of crashes for the \( i \)th observation. \( X_i \) denotes the vector of explanatory factors and \( \theta \) the vector of corresponding coefficients. \( \alpha \) is the over-dispersion parameter.

A hierarchical binomial-logistic model is employed to estimate crash severity. Given that each crash occurrence is considered a Bernoulli trial, the crash outcome with respect to injury severity should follow a binomial distribution. The number of KSI crashes \( k_{i}^{\text{KSI}} \) follows the binomial distribution defined as
\[
k_{i}^{\text{KSI}} \sim \text{Binomial}(x_{i}^{\text{KSI}}, y_{i}),
\]  
(2)
where \( y_{i} \) is the total crash risk for the \( i \)th observation and \( x_{i}^{\text{KSI}} \) is the probability of KSI crashes.

Therefore, the probability function of \( k_{i}^{\text{KSI}} \), conditional on \( y_{i} \) total crashes, can be defined as
\[
P(k_{i}^{\text{KSI}} | y_{i}) = \binom{y_{i}}{k_{i}^{\text{KSI}}} \left( x_{i}^{\text{KSI}} \right)^{k_{i}^{\text{KSI}}} \left( 1 - x_{i}^{\text{KSI}} \right)^{y_{i} - k_{i}^{\text{KSI}}}
\]  
(3)
Here, we can establish a logit function to model the associations between the binomial probability and possible factors defined as
\[
\text{Logit}(x_{i}^{\text{KSI}}) = \beta_{i}^{\text{KSI}} X_i,
\]  
(4)
where \( \beta_{i}^{\text{KSI}} \) is the vector of parameters measuring the correlations between explanatory factors and crash severity.

Hence, we have
\[
P(k_{i}^{\text{KSI}} | y_{i}) = \binom{y_{i}}{k_{i}^{\text{KSI}}} \left[ \frac{\exp(\beta_{i}^{\text{KSI}} X_i)}{1 + \exp(\beta_{i}^{\text{KSI}} X_i)} \right]^{k_{i}^{\text{KSI}}} \left[ \frac{1}{1 + \exp(\beta_{i}^{\text{KSI}} X_i)} \right]^{y_{i} - k_{i}^{\text{KSI}}}
\]  
(5)
Consequently, the joint probability function for crash occurrence and crash severity can be obtained by combining Eqs. (1) and (5) as
\[
P(y_i, k_{i}^{\text{KSI}}) = P(y_i) \times P(k_{i}^{\text{KSI}} | y_{i})
\]
\[
= \frac{\Gamma[(1/\alpha) + y_i]}{\Gamma(1/\alpha) y_i!} \left( \frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left( \frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{y_i}
\times \binom{y_{i}}{k_{i}^{\text{KSI}}} \left[ \frac{\exp(\log(\text{exposure}) + \theta X_i)}{1 + \exp(\log(\text{exposure}) + \theta X_i)} \right]^{k_{i}^{\text{KSI}}} \left[ \frac{1}{1 + \exp(\log(\text{exposure}) + \theta X_i)} \right]^{y_{i} - k_{i}^{\text{KSI}}}
\]  
(6)

3.2. Model estimation

The full Bayesian method, using a Markov chain Monte Carlo (MCMC) approach, is applied to the proposed joint probability model. The full Bayesian method treats all unknown parameters as random variables under a prior distribution. Instead of giving point estimates, as in the classical maximum likelihood approach, the full Bayesian method uses the posterior distribution of the parameters to summarize the results in the proposed model (Gelman et al., 2004).

According to Bayes’ theorem, by integrating the prior distribution of parameters and the likelihood function, the posterior distribution of parameters can be estimated by the following function:
\[
f(\theta | y) = \frac{f(y | \theta)f(\theta)}{f(y)} \propto f(y | \theta)f(\theta)
\]  
(7)
where \( y \) is the observed outcome and \( \theta \) is the parameter estimate. Therefore, the likelihood function can be defined as
\[
f(y | \theta) = \prod_{i=1}^{n} f(y_i | \theta)
\]  
(8)
Appropriate specification of the prior distribution \( f(\theta) \) is a critical element of Bayesian inference. For instance, it is essential to have prior information on and a good understanding of the variables of interest. If prior information for particular parameters is not available, we can set up a diffuse prior distribution as
\[
\theta \sim \text{Normal}(0, \sigma^2 I_m),
\]  
(9)
where \( \sigma^2 \) is very large, \( I \) is an identity matrix of \( m \) dimensions, and \( m \) equals the number of parameters.

In a full Bayesian framework, the inferences are determined by the Markov chain Monte Carlo (MCMC) simulation approach that draws the sample from the prior distribution of unknown parameters in an iterative procedure. Such a process should be repeated until the distribution of parameter estimates converges and gives the posterior distribution. The Metropolis–Hastings algorithm and Gibbs sampling are two common sampling approaches that follow the MCMC method. Gibbs sampling can be regarded as a special case of the Metropolis–Hastings sampling algorithm, in which a candidate estimate is always accepted rather than being accepted at a particular probability. In addition, the Gibbs sampling approach is capable of drawing a sample from the distributions of parameters other than that of the variable of concern, thus ensuring that the estimates will converge more easily. This quality has made the widely used Gibbs sampling method a ‘golden standard’. Yet Gibbs sampling might not be suitable when the form of the parameter distribution is sophisticated (Ntzoufras, 2009).

Possible ways of evaluating convergence of the estimates in the iterative procedure include estimating the Markov chain error, plotting the autocorrelation distributions, plotting the generated sample values, and computing Gelman–Rubin statistics for multiple chain simulation (Spiegelhalter et al., 2003).

The validity of the model can be evaluated by comparing the observed and replicated data. The replicated data can be predicted in the simulation process set out by the posterior predictive distribution. For instance, a chi-square test statistic can be calculated by
\[
\chi^2(y, \theta) = \sum_{i=1}^{n} \frac{[y_i - E(y_i | \theta)]^2}{\text{Var}(y_i | \theta)}.
\]  
(10)
where \( y_i \) denotes either the observed response or the replicated response. In each iteration \( t \) of an MCMC simulation, the difference between \( \chi^2(y^{\text{rep}}, \theta^{(t)}) \) and \( \chi^2(y, \theta^{(t)}) \) is monitored, as is the corresponding posterior predictive \( p \)-value. The results indicate that the proposed model does not fit well if the \( p \)-value is close to 0 or 1. Gelman et al. (2004) suggests that major failures should be considered if the \( p \)-value is less than 0.01 or more than 0.99.

4. Results

This study establishes two joint probability models controlling for distance and time exposure, respectively, to measure the relationship between speed, other possible risk factors, and crash occurrence and severity simultaneously. In the proposed full Bayesian modeling framework with the MCMC approach, three
chains of 30,000 iterations are adopted in each simulation process. All posterior estimates are within the range for not producing strong periodicities and tendencies, as indicated by the trace plots. In addition, Gelman–Rubin convergence diagnostic results indicate that both the inter-sample and intra-sample variabilities are stable and close to one. In other words, the two proposed models both converge. Table 2 reports the results of the two proposed joint probability models. To assess goodness-of-fit, the p-values of the chi-square statistics are calculated as 0.12 and 0.09 for the models controlling for distance and time exposure, respectively. The proposed models generally fit well with the observations.

As shown in Table 2, there are no major differences in the parameter estimates between the model controlling for distance and that controlling for time exposure, other than for average speed. For instance, when distance exposure is controlled for, average speed [Mean = 0.02; 95% Cls = (-0.03, -0.01)] is negatively related to the likelihood of a crash occurring at the 5% level of significance. In contrast, when controlling for time exposure, average speed [Mean = 0.01; 95% Cls = (0.00, 0.01)] is positively related to the likelihood of a crash occurring at the 5% significance level.

For crash occurrence, rainfall is positively related to the likelihood of a crash occurring, controlling for both distance [Mean = 0.02; 95% Cls = (0.01, 0.03)] and time exposure [Mean = 0.02; 95% Cls = (0.01, 0.03)], at the 5% significance level. Furthermore, lane changing opportunities is positively associated with the likelihood of a crash occurring, controlling for both distance exposure [Mean = 0.20; 95% Cls = (0.12, 0.28)] and time exposure [Mean = 0.18; 95% Cls = (0.10, 0.26)], at the 5% significance level. The road segment with central divider is related to lower likelihood of a crash occurring, controlling for both distance exposure [Mean = -0.55; 95% Cls = (-1.06, -0.09)] and time exposure [Mean = -0.54; 95% Cls = (-1.02, -0.06)], at the 5% significance level. In addition, for the road segment with bus stop, the probability of a crash occurring is higher, controlling for both distance exposure [Mean = 0.41; 95% Cls = (0.13, 0.68)] and time exposure [Mean = 0.45; 95% Cls = (0.18, 0.73)], at the 5% significance level. Moreover, the crash risk on Tuesdays is lower than that on Sundays or public holidays, controlling for both distance exposure [Mean = -0.52; 95% Cls = (-1.00, -0.06)] and time exposure [Mean = -0.52; 95% Cls = (-0.99, -0.04)], at the 5% significance level.

Table 2 also shows that for crash severity, average speed is positively related to the likelihood of a KSI crash, controlling for both distance exposure [Mean = 0.03; 95% Cls = (0.01, 0.06)] and time exposure [Mean = 0.03; 95% Cls = (0.01, 0.06)], at the 5% significance level. Rainfall is negatively related to the likelihood of a KSI crash, controlling for both distance exposure [Mean = -0.09; 95% Cls = (-0.19, -0.01)] and time exposure [Mean = -0.08; 95% Cls = (-0.18, -0.01)], at the 5% significance level. Furthermore, the number of diverging ramps is also negatively related to the likelihood of a KSI crash, controlling for both distance exposure [Mean = -0.22; 95% Cls = (-0.43, -0.02)] and time exposure [Mean = -0.21; 95% Cls = (-0.41, -0.01)], at the 5% significance level.

5. Discussion

We establish two joint probability models to reveal the association between speed, other possible factors, and crash occurrence and crash severity simultaneously. We also investigate differences in association measures when controlling for different exposure measures.

First, the relationship between speed and the likelihood of a crash occurring differs according to which exposure measure is adopted. For instance, the results indicate that crash risk is negatively associated with average speed when controlling for distance exposure. Although it is not uncommon for distance exposure to be controlled for when modeling the association between crash...
risk and speed, results from prior studies were inconclusive on this point (Solomon, 1964; Shinar, 1998; McCarthy, 1998; Aarts and van Schagen, 2006; Garber and Gadirau, 1988). Some researchers argue that roadway segments designed for higher speeds are usually well planned and developed, and should deliver better road safety performance. Therefore, crash frequency should decrease with speed. The current study reveals not only a negative relationship between speed and crash risk (controlling for distance exposure), but also the possible influences of explanatory factors including road design, weather conditions, and temporal distribution on respective relationship. Nevertheless, information on other possible factors related to crash occurrence, including traffic composition and driver behavior, was not available for this study. These factors are worthy of exploration in future research.

Notwithstanding the foregoing finding, there is also an argument that this negative relationship between speed and crash occurrence, observed when controlling for distance exposure, may distort the role of speed, because the likelihood of a vehicle traveling at a higher speed being involved in a crash is underestimated for a given travel distance due to the shorter time of exposure (Chipman et al., 1992). This issue has received less attention because travel time and speed data are generally unavailable for accident analyses. Given the availability of speed data for this study, we are able to investigate this problem by controlling for time exposure on highways. Our analysis finds that crash risk is positively associated with average speed when time exposure is controlled for. This helps to shed light on the interesting problem of whether higher speed is related to a higher crash risk; our results indicate that given the same time duration, higher vehicle speed would be associated with higher crash risk when correcting for the reduction in time exposure per unit of distance traveled. This means that high-speed roads may not necessarily be safer than lower-speed roads when the time exposure measure is properly accounted for.

Turning to injury severity, it is apparent that both models examined here point to a similar observation whereby crash severity is positively related to speed, controlling for both distance and time exposure. This finding is consistent with that of a previous study (Hauer, 2009), which argues that as the amount of energy released is greater when the impact speed is higher, the harm done and the likelihood of mortality and serious injury are greater.

The implications of the relationship between speed, crash occurrence and crash severity are illustrated in Fig. 2. For instance, the relationships between average speed and (i) total crash risk, (ii) KSI crash risk, and (iii) the risk of a crash causing slight injury are
evaluated. As shown in Fig. 2(a), when controlling for distance exposure, total crash risk decreases with speed, whereas the risk of a KSI crash increases with speed.

As shown in Fig. 2(b), when controlling for time exposure, both total crash risk and KSI crash risk increase with speed. In particular, the risk of a KSI crash increases at a rate remarkably higher than that of total crash risk because the likelihood of a KSI crash is more sensitive to speed. Moreover, when speed increases beyond 65 km/h, the risk of a crash causing slight injury begins to decline. When speed is around 110 km/h, the risk of a KSI crash is equal to that of a slight injury crash. Furthermore, when speed increases beyond 110 km/h, the risk of a KSI crash is much higher than that of a slight injury crash.

Speed dispersion was also expected to be highly related to crash risk. However, in the proposed model, there is no evidence that the standard deviation of speed is significantly associated with the likelihood of crash occurrence or crash severity. However, it is worth noting that the speed dispersion used in this study does not represent the variations in speed seen in a mixed traffic stream with different types of vehicles, but is a proxy for the variability of traffic conditions experienced by drivers in a spatio-temporal domain on the highway.

Since the Power Model has been widely adopted to connect speed and road safety, we attempt to apply the Power Model to our model by incorporating logarithmically transformed speed into the link function. For the crash occurrence (i.e., injured crash) model, the exponent estimate of speed is 0.50. Such an estimate is much lower than that of Nilsson’s model (Nilsson, 2004), but is close to that (i.e., 1.2, 95% CI [0.7, 1.7]) of Cameron and Elvik’s model (2010) for urban/residential roads. For the crash severity (i.e., KSI crash) model, the exponent estimate of speed is 1.44. Such an estimate is close to that of Nilsson’s model (i.e., Exponent of KSI crash minus Exponent of Casualty Crash = 3 minus 2 = 1). There are two possible explanations for why the results of the current model deviate from those of Nilsson’s model. First, our model measured the risk of crash and crash severity on urban roads, whereas Nilsson’s model measured that of rural highways and freeways. Additionally, our model measured the average speed over a very short time window (4 h), whereas Nilsson’s model measured the average speed over a longer period of time. In future, it is worth exploring the differences in the effects of speed on crash risk with respect to different road types.

For weather conditions, the results indicate that rainfall is related to the likelihood of a crash occurring with a positive sign, but to the likelihood of a KSI crash with a negative sign. This may be because rain reduces skidding resistance and driver visibility, and thus increases the overall crash risk. However, the impact force exerted during collisions may be reduced as driver alertness may increase when driving in rainy conditions, and crashes in such weather conditions are, therefore, possibly less severe (Hermans et al., 2006; Khattak et al., 1998; Edwards, 1998; Fridstrom et al., 1995; Sze and Wong, 2007).

For road design, increased lane-changing opportunities are found to be related to an increase in the likelihood of a crash occurring. This result could be attributable to increased vehicle interaction due to the increase in the number of permissible lane-cutting and overtaking activities, thereby raising the incidence of traffic conflict and collisions. Similarly, the presence of a bus stop is related to a higher crash risk possibly due to frequent pedestrian activity around the bus stop, leading to a higher risk of vehicle-pedestrian conflict. Indeed, of the crashes involving pedestrians in the observed road segments, 93% occurred on roads with bus stops. Moreover, there may be potential for more conflict between buses entering or leaving bus bays and other vehicles in the main traffic stream. In addition, our prediction that the presence of a central divider would be related to a lower risk of a crash occurring is confirmed, as the opposing traffic flows are separated. Moreover, an increase in the number of diverging ramps is related to the reduction in the risk of KSI crash. This may be because the increase in road space for defensive driving behavior at the area near diverging ramps leads to a reduction in the impact of the collision in the event of a crash.

With respect to temporal distribution, crash risk on Tuesdays is found to be lower than that on Sundays or public holidays. Although it may not be possible to elaborate explicitly on the variation in crash risk across different time periods, the influences of unidentified factors that vary over time are controlled for in the proposed model, including both time of day and day of the week.

6. Conclusions

Based on comprehensive crash, traffic flow, and speed data for Hong Kong’s roadway network, we investigate the effects of different forms of exposure measures, i.e., time and distance exposure, on the relationship between speed, other possible risk factors and crash risk. Taking advantage of detailed speed data obtained from GPS probes, it is possible to reveal the relationship between speed and crash risk when controlling for time exposure. A joint probability model is applied to analyze crash occurrence and crash severity simultaneously under the MCMC approach full Bayesian framework.

The results indicate the positive relationship between speed and the risk of a crash occurring when controlling for the effect of time exposure. This contrasts with the results obtained when controlling for distance exposure, which indicates the opposite relationship. This finding helps shed light on the interesting argument that controlling for distance exposure may distort the overall relationship between crash risk and speed, because of the lower time exposure per unit of distance traveled on high-speed roads. From the observations made in this study, it is apparent that high-speed roads may not necessarily be safer than lower-speed roads when a proper exposure measure is used. However, the results for crash severity show that injury severity is generally positively associated with speed, irrespective of whether distance or time exposure is used.

Our results do not provide evidence suggesting that speed dispersion, as a proxy for the variability of traffic conditions in the spatio-temporal domain of a highway, is significantly related to either crash risk or crash severity. However, it would be worthwhile exploring variations in the associations between speed, speed dispersion, and crash risk across different crash types in future research.

The joint probability approach adopted in this study integrates both types of risk assessed in crash models—the risk of an accident occurring given a certain level of exposure and the risk of injury severity in the event of an accident—into a unified analysis that gives us a better understanding of the roles played by exposure and speed in the likelihood of traffic accidents occurring and the severity of injuries suffered by casualties. This approach supports the holistic assessment of traffic accidents and helps to identify the relationship between crash risk and possible risk factors. Moreover, for the latter case in which a risk factor is related to both types of risk, the analysis helps to give an overall picture of the risk tradeoff between accident occurrence and injury severity for this factor.

Acknowledgments

The work described in this paper was partially supported by a Research Postgraduate Studentship, an Outstanding Researcher Award, the Engineering Postdoctoral Fellowship Programme of the
University of Hong Kong, a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU7176/07E), and a grant from National Natural Science Foundation of China (Grant No. 61021063). We would also like to thank Concord Pacific Satellite Technologies Limited and Motion Power Media Limited for providing the GPS taxi data.

References


Khattak, A., Kantor, P., Council, F., 1998. Role of adverse weather in key crash types on limited-access roadways: implications for advanced weather systems. Transportation Research Record 1621, 10–19.


